

DEVELOPMENT OF A FAIR BASED APPROACH TO DATA MANAGEMENT AND SHARING WITHIN EUROFUSION

P. I. Strand
Chalmers University of Technology
Göteborg, Sweden
Par.Strand@chalmers.se

On behalf of
EUROfusion Working Group on Open Data
Fair for fusion consortium

Abstract

The European fusion research activities have over the last decades generated a vast and varied set of data. Even if a survey of the generated data is restricted to EUROfusion, the implementation of the fusion research programme under the EU Horizon 2020 framework programme, the volume and diversity of the data that need to be catalogued and reviewed make the task of organizing and making the data available very challenging. Nevertheless, there are strong scientific drivers as well as incentives and mandates from national research agencies suggesting that a more coherent approach to data dissemination and sharing would be very necessary for the fusion research community. In anticipation of stricter requirements on data management for the EUROfusion implementation within the 2021-2025 EURATOM programme both policies and technical implementations to facilitate a partial transition to open and FAIR (Findable, Accessible, Interoperable, and Reusable) data sharing are being developed within the continued European activities, guided by the principle “as open as possible, as closed as necessary”.

Here, the progress and status of the work done in two related activities are reported: a) the EUROfusion Open Data Working Group and b) the Fair for Fusion project (funded under EURATOM). Together they cover the necessary and needed policy and technological elements to formulate a data management plan and propose a corresponding implementation plan. Restrictions on the scope of data being made available are being addressed, policies related to data access, data validation and embargo periods are discussed and the technical solutions proposed in relation to metadata description, persistent identifiers, data ingestion, and data access and exchange formats (IMAS based) as well as authentication and authorization are being motivated.

The proposed infrastructure has been developed based on several use cases covering both experimental and modelling data from different user perspectives: active, researchers, managers, and the general public. The researchers’ point of view had a strong focus on multidevice access and the ability to do both searches on global quantities for the full range of supported devices and to make bespoke data collections based on a range of criteria covering both single and multiple devices. The proposed infrastructure lends itself well to data centric analytics work and these aspects are introduced and discussed in view of prototypes and demonstrators that have been designed

